Appendix

Appendix A1 Study characteristics: Sivin-Kachala & Bialo, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Sivin-Kachala, J., & Bialo, E. (2005). Fluency Formula second grade study, Long Island, New York 2003-2004: Evaluation research on the effectiveness of Fluency Formula. Retrieved from Scholastic Education Web site: http://teacher.scholastic.com/products/fluencyformula/pdfs/FF_EffectivessReport.pdf
Participants	The study began with 252 second-grade students (127 in the intervention condition) from 12 classrooms (six classrooms in each condition). Participants came from three schools in one district and two schools in a second district (personal communication). Classrooms were matched on students' reading ability, demographic variables, and teacher characteristics and then randomly assigned to study conditions. Because *Fluency Formula™* does not target high-performing students, analyses were limited to the 143 students scoring at or below the 75th percentile on the Edformation Oral Fluency Assessment (OFA) pretest. Fifteen additional students were removed from analyses: one student who moved from the comparison to intervention group, eight students who exited the program, four students who missed at least half the intervention lessons, and two students with a large number of absences. The final analyses were based on data from 66 intervention students and 62 comparison students. According to data provided in the study, the remaining students (analysis sample) from the intervention and comparison groups were comparable on pretest achievement measures (OFA).[1]
Setting	The study was conducted in five schools[2] across two suburban school districts in the northeastern United States.
Intervention	*Fluency Formula™* was delivered as a supplement to the participating schools' standard reading/English language arts curriculum. Intervention group teachers followed the six-unit curriculum sequence in the *Fluency Formula™ Professional Guide*. Students classified as "low initial ability" (based on the pretest of oral fluency) received four days a week of *Fluency Formula™* instruction (two days of whole-class instruction plus two days of small-group instruction). Students classified as "high initial ability" received two days of whole-class *Fluency Formula™* instruction. Once a week, students received a 15-minute take-home assignment.
Comparison	The comparison group students received the participating schools' standard reading/English language arts curriculum with no supplemental materials or instruction. The comparison group teachers were not exposed to the *Fluency Formula™* materials or professional development.
Primary outcomes and measurement	The primary outcome measure in the fluency domain is the Edformation Oral Fluency Assessment (OFA) and the primary outcome in the comprehension domain is the Woodcock-Johnson III Tests of Achievement: Passage Comprehension subtest (see Appendixes A2.1 and A2.2 for more detailed descriptions of outcome measures).[3]
Teacher training	Intervention group teachers attended one after-school professional development session (about two and a half hours) that presented the theoretical basis, components, and implementation of the program. Teachers followed the instructional sequence detailed in the *Fluency Formula™ Professional Guide*.

1. The authors also used the Woodcock-Johnson III Basic Reading cluster score as a pretest and indicated that there were no statistically significant differences between the treatment and comparison students. These data were provided by the study authors in personal communication. For one outcome (comprehension), the authors analyzed the two (unidentified) districts separately. In that analysis, the comparison group in one of the districts scored significantly lower than the intervention group on pretest. The WWC, however, considered the pooled data, rather than data for the separate districts.
2. This information was provided by the study authors in personal communication.
3. The study also included student outcomes in the alphabetics domain, but those outcomes were not reviewed for rating purposes because the data were not reported in the study.

Appendix A2.1 Outcome measures in the fluency domain

Characteristic	Description
Edformation Oral Fluency Assessment (OFA)	This test measures the number of words correct per minute (WCPM) that students read using three brief grade-level passages (200-word minimum). These passages include both fiction and nonfiction text. The norms for this test are updated by Edformation each school year[1] (as cited in Sivin-Kachala & Bialo, 2005).

1. The study authors reported both raw and normal curve equivalent (NCE) scores. The WWC used the raw score results.

Appendix A2.2 Outcome measures in the comprehension domain

Characteristic	Description
Woodcock-Johnson III Tests of Achievement: Passage Comprehension subtest	This standardized test measures children's ability to match words and pictures of objects the words stand for and to identify missing key words in the reading passage (as cited in Sivin-Kachala & Bialo, 2005).

Appendix A3.1 Summary of study findings included in the rating for the fluency domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome[2] (standard deviation[3]) — *Fluency Formula™ (FF)* group	Authors' findings from the study: Mean outcome[2] (standard deviation[3]) — Comparison group	WWC calculations: Mean difference[4] (*FF* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Sivin-Kachala & Bialo, 2005 (randomized controlled trial)[8]								
Oral Fluency Assessment (OFA)	Grade 2	12/128	84.85 (19.65)	78.62 (27.88)	6.24	0.26	ns	+10
Domain average[9] for fluency (Sivin-Kachala & Bialo, 2005)						0.26	ns	+10

ns = not statistically significant

1. This appendix reports the findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same study are not included in these ratings, but are reported in Appendix A4.
2. The means for the intervention and comparison groups were calculated by the WWC by combining initial fluency ability subgroup means (see Technical Details of WWC-Conducted Computations). The reported intervention group mean equals the comparison group mean plus the mean difference.
3. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. Standard deviations for the combined intervention and comparison groups were calculated by the WWC based on subgroup standard deviations.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference was calculated using the difference-in-difference (gain scores) approach; see Technical Details of WWC-Conducted Computations.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Sivin-Kachala & Bialo (2005), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome[2] (standard deviation[3]) — *Fluency Formula™ (FF)* group	Authors' findings from the study: Mean outcome[2] (standard deviation[3]) — Comparison group	WWC calculations: Mean difference[4] (*FF* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Sivin-Kachala & Bialo, 2005 (randomized controlled trial)[8]								
Woodcock-Johnson III: Passage Comprehension subtest	Grade 2	12/128	477.46 (10.88)	480.82 (13.05)	–3.36	–0.28	ns	–11
Domain average[9] for comprehension (Sivin-Kachala & Bialo, 2005)						–0.28	ns	–11

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The reported intervention group mean equals the comparison group mean plus the mean difference.
3. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. Standard deviations for the combined intervention and comparison groups were calculated by the WWC based on subgroup standard deviations from two districts.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference was calculated using the difference-in-difference (gain scores) approach; see Technical Details of WWC-Conducted Computations.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Sivin-Kachala & Bialo (2005), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 Summary of subgroup findings for the fluency domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome[2] (standard deviation[3]) — *Fluency Formula™ (FF)* group	Authors' findings from the study: Mean outcome[2] (standard deviation[3]) — Comparison group	WWC calculations: Mean difference[4] (*FF* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Sivin-Kachala & Bialo, 2005 (randomized controlled trial)—Low-ability subgroup[8]								
Oral Fluency Assessment (OFA)	Grade 2	12/72	74.28 (17.53)	60.83 (22.40)	13.45	0.66	ns	+25
Sivin-Kachala & Bialo, 2005 (randomized controlled trial)—High-ability subgroup[8]								
Oral Fluency Assessment (OFA)	Grade 2	12/56	98.63 (14.43)	101.67 (13.79)	–3.04	–0.21	ns	–11

ns = not statistically significant

1. This appendix presents ability subgroup findings for measures that fall in the fluency domain. Total group scores were used for rating purposes and are presented in Appendix A3.1.
2. The intervention group mean equals the comparison group mean plus the mean difference.
3. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference was calculated using the difference in difference (gain scores) approach; see Technical Details of WWC-Conducted Computations.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Sivin-Kachala & Bialo (2005), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A5.1 *Fluency Formula™* rating for the fluency domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of fluency, the WWC rated *Fluency Formula*™ as having potentially positive effects. It did not meet the criteria for the positive effects because there was only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Fluency Formula*™ was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. The one study reviewed showed substantively important positive effects.

and

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No study showed statistically significant or substantively important negative effects or indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study met WWC evidence standards.

and

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Fluency Formula™* rating for the comprehension domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of comprehension, the WWC rated *Fluency Formula*™ as having potentially negative effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, no discernible effects, and negative effects) because the single study that met WWC standards showed substantively important negative effects.

Rating received

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Met. The one study reviewed showed substantively important negative effects.

and

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed statistically significant or substantively important positive effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study met WWC evidence standards.

and

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Not met. One study showed substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed statistically significant or substantively important positive effects.

and

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. One study showed substantively important negative effects.

(continued)

Appendix A5.2 *Fluency Formula™* rating for the comprehension domain *(continued)*

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed statistically significant or substantively important positive effects.

or

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. One study showed substantively important negative effects and no study showed indeterminate effects.

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Not met. One study showed substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence rating by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	0	0	0	na
Comprehension	1	5	128	Small
Fluency	1	5	128	Small
General reading achievement	0	0	0	na

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."